Jarrod Kahn garrod.kahn@gmail.com

EXPERIENCE

Titles reflect positions held at the time of departure.

Meta San Francisco Bay Area, CA

Research Engineer Nov. 2024 - Present

Google San Francisco Bay Area, CA

Senior Staff Research Engineer at Google DeepMind Apr. 2023 – Nov. 2024

I'm the lead of the Gemini/LaMDA Research Tool Use Group. I created the modeling techniques for tool use that drive generative AI applications across Alphabet, including Gemini, Magi, and Google Cloud.

Staff Software Engineer at YouTube

Tech lead and manager of YouTube's Home Engagement Modeling team. My team is a high velocity, high impact group applying advanced ML techniques at an enormous scale.

- Model scale: I led a multiyear effort to scale our models by a factor of 1,000 using a large fleet of 4th generation TPU training and serving hardware. This effort resulted in several unprecendented increases in daily active users, which led to the broader shift in strategic investment towards scaling models across the entire organization.
- **Knowledge distillation**: My team was the first at YouTube to deploy large model distillation. Model training and serving efficiency has become a top priority across all of YouTube, and knowledge distillation is a key strategic initiative to improve scaling economics.
- Training infrastructure: I led the effort to develop the training framework used by all of YouTube for models trained on user interaction logs. The framework allows engineers to write TensorFlow or JAX and has first class support for TPUs. It emphasizes ease of development, testing, code over configuration, and code sharing.

 Today there are over 30 distinct model types and thousands of models using this framework.
- Modeling techniques: Under my technical leadership, the team has applied many innovations to YouTube models, including attention based user modeling, AutoML, quantile regression, dense cross networks, uncertainty estimation, bias correction and many more.

VevoSan Francisco, CASenior Machine Learning EngineerDec. 2015 – Jul. 2018

One of the first engineers on Vevo's recommender systems team. Collect and analyze data from the platform and train models that power personalization of music video applications.

- Recommender Systems: Developed, trained, and deployed models to recommend items to users.
- Entity Modeling: Modeled users and videos using large, noisy, and sparse data in a semantic embedding space.
- Transfer Learning: Train models using these semtantic embeddings & past behavior to predict user preference.
- ML Infrastructure: Design, implement, and deploy data collection, model training, and delivery systems.

Yola San Francisco, CA
Lead Software Engineer Jul. 2013 – Dec. 2015

Forge Aerospace Boulder, CO

Part Time Software Engineer

Jan. 2013 – Jul. 2013

University of Colorado

Boulder, CO

Part Time Web Developer Jan. 2011 – Aug. 2012

EDUCATION

University of Southern California

Master of Science in Computer Science

University of Colorado

Boulder, CO

Los Angeles, CA

Jul. 2018 - Apr. 2023

Bachelor of Science in Computer Science Bachelor of Arts in Economics